In Silico Analysis and SNP identification for GBSS-I and branching enzyme genes involved in Starch metabolism in Rice cultivars

Enrollment No: 091519, 091722

Name of Students: Pankaj Sharma and Jessica Sharma

Name of supervisor: Dr. Harvinder Singh

May – 2013

Submitted in partial fulfillment of the Degree of

4 Year Degree Programme B. Tech

DEPARTMENT OF BIOTECHNOLOGY

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,

WAKNAGHAT

TABLE OF CONTENTS

CERTIFICATE

This is to certify that the work titled "***In Silico* Analysis and SNP identification for GBSS-I and branching enzyme genes involved in Starch metabolism in Rice cultivars**" submitted by "**Pankaj Sharma** and **Jessica Sharma**" in partial fulfillment for the award of degree of 4 Year Degree Program B.Tech of Jaypee University of Information Technology, Waknaghat has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor .....

Name of Supervisor: Dr. Harvinder Singh

Designation: Assistant Professor

Date 25/5/13

(II)

ACKNOWLEDGMENT

We take this opportunity to express our profound gratitude and deep regards to our guide **Dr. Harvinder Singh** for his exemplary guidance, monitoring and constant encouragement throughout the course of this thesis. The blessing, help and guidance given by him time to time and the support of the parents shall carry us a long way in the journey of life on which we are about to embark.

We owe my special debt of gratitude and heart full thanks to **Mr. Jibesh Kr. Padhan** for devoting his precious time and guiding us with his valuable suggestions and for bearing our mistakes and shortcomings.

Signature of the student ..... Pankaj Sharma ...and Jessica Sharma

Name of Student Pankaj Sharma and Jessica Sharma

Date ...25/05/2013...

(III)

ABSTRACT

Starch is a major component of human diets. The relative contribution of variation in the genes of starch biosynthesis to the nutritional and functional properties of the rice was evaluated in a rice breeding population. Sequencing of GBSS 1 gene involved in starch synthesis in a population of 17 rice breeding lines resulted in the discovery of putative 29 SNPs out of which 16 were in intronic region and 13 functional SNPs were observed in exonic region. Out of 13 functional SNPs observed 31% were Synonymous and 45% were Non Synonymous. The latter are beneficial to our study as they result in variation among the species. The results indicated the possibility of developing high palatability cultivars through molecular breeding for key genes related to *indica* rice eating quality, involved in starch biosynthesis. Thus, this study supports the previous findings that the GBSS-I gene affects amylose content. A further study on corerelation between the SNP changes and the amylose content is proposed to identify the improved variety of rice.

Pankaj Sharma and Jessica Sharma

Signature of Students

Name: Pankaj Sharma and Jessica Sharma

Date 25/05/2013

Signature of Supervisor

Name: Dr. Harvinder Singh

Date 25/5/13

LIST OF TABLES AND FIGURES

Chapter 1

Section 1.1 – Introduction

Oryza sativa, commonly known as Asian rice, is the plant species most commonly referred to in English as rice. *Oryza sativa* is the cereal with the smallest genome, consisting of just 430Mb across 12 chromosomes. It is renowned for being easy to genetically modify, and is a model organism for cereal biology.(CECAP, Phil Rice and IIRR. 2000")

Rice is a major human food composed largely of starch. Starch properties determine the key functional properties of rice such as cooking temperature and influence human health through its contribution to the glycemic index and levels of resistant starch. The incomplete digestion-absorption of resistant starch in the small intestine leads to non-digestible starch fractions with physiological functions similar to dietary fibre with significant beneficial impacts. (Oka (1988))

Oryza sativa is a kind of grass, which grows best when submerged in water. It grows in pland areas, irrigated areas; rain fed lowland areas, and flood-prone areas. Rice is highly adaptable and can be grown even in diverse environments. It resembles a weed, 2 to 5 feet tall, depending on the variety and depth of submersion. It has round, hollow, jointed stems, rather flat, sessile leaf blades, and a terminal panicle. The grain is produced on nodding panicles of spikelets. It looks like a smooth glistening ovoid particle, emerald green in color (during ripening stage, however, it turns golden yellow).

The origins of rice have been debated for some time, but the plant is of such antiquity that the precise time and place of its first development will perhaps never be known. It is certain, however, that the domestication of rice ranks as one of the most important developments in history, for rice is the longest, continuously grown cereal crop in the world. Botanical and linguistic evidence point to the early origin of domesticated rice along a broad arc from eastern India through Myanmar, Thailand, Laos, Northern Vietnam, and into southern China. The earliest and most convincing evidence for domestication of rice in Southeast Asia was discovered in 1966 at Non NokTha in the Korat area of Thailand. These remains have been confirmed as dating from at least 4000 B.C.

Oryza sativa was the cereal selected to be sequenced as a priority and has gained the status "model organism". It has the smallest genome of all the cereals: 430 million nucleotides and it can serve as a model genome for one of the two main groups of flowering plants, the monocotyledons. Because it has been the subject of studies on yield, hybrid vigor, genetic resistance to disease and adaptive responses, scientists have taken advantage of the existence

of a multitude of varieties that have adapted to a very wide range of environmental conditions, Rice production represents 30% of the world cereal production today. It has doubled in the last 30 years, in part due to the introduction of new varieties, but its present growth barely follows consumption: in 2025 there will be 4.6 billion people that depend on rice for their daily nourishment, compared with three billion today. A new leap in production is therefore expected. At the same time, small producers will have to use land which is less favorable for cultivation, such as brackish or briny soils, and the availability of water resources will become more and more problematic.

Two research teams have sequenced related subspecies of rice. Essential biological information from the rice genome will undoubtedly improve our understanding of the basic genomics and genetics of other related and economically significant crops, not only wheat, corn, sorghum, and members of the grass family, but also dicot crops such as soybean and cotton. The finished genomes will take some of the guesswork out of plant breeding as breeders will be able to determine whether a seed contains a particular gene through genetic analysis. If a gene is known to contribute to trait of interest, variants of this gene can be examined in other varieties.

Section 1.2 – Classification

There are about 120,000 varieties known to exist. Two of the types sequenced are Indica and Japonica, the japonica varieties have narrow dark green leaves, medium-height tillers, and short to intermediate plant height. It is usually grown in cooler subtropics and temperate climates, such as Japan, Portugal, Spain, USSR, Italy, and France. The traditional indica rice varieties, widely grown throughout the tropics and subtropics, are tall and heavy tillering with long, narrow, light green leaves.

1. ***Japonica* varieties are** usually cultivated in dry fields, in temperate East Asia, upland areas of Southeast Asia and high elevations in South Asia.

2. ***Indica* varieties are** mainly lowland rices, grown mostly submerged, throughout tropical Asia. Rice is known to come in a variety of colors, including: white, brown, black, purple, and red.

3. A third subspecies, which is broad-grained and thrives under tropical conditions, was

Identified based on morphology and initially called *javanica*, but is now known as *tropical japonica*. Examples of this variety include the medium grain 'Tinawon' and 'Unoy' cultivars, which are grown in the high-elevation rice terraces of the Cordillera Mountains of northern Luzon, Philippines.

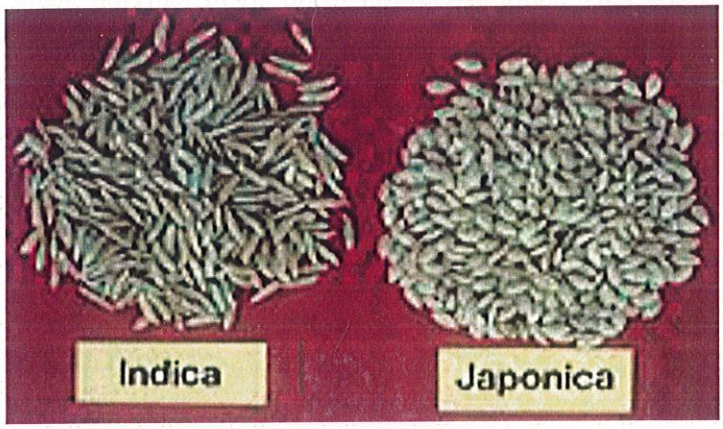


Fig1: Rice Classification

Kingdom	Plantae
Class	Monocots
Family	Poaceae
Species	*O.sativa*
Division	Angiosperms
Order	Poales
Genus	Oryza

Table 1: Classification of Oryza sativa (Glaszmann, J. C. (2004))

Section 1.3 – Starch Metabolism

Starch oramylumis a carbohydrate consisting of a large number of glucose units joined by glycosidic bonds. This polysaccharide is produced by all green plants as an energy store. It is the most common carbohydrate in the human diet and is contained in large amounts in such staple foods as potatoes, wheat, maize (corn), rice, and cassava.

Pure starch is a white, tasteless and odorless powder that is insoluble in cold water or alcohol. It consists of two types of molecules: the linear and helical amylose and the branched amylopectin. Depending on the plant, starch generally contains 20 to 25%

amylose and 75 to 80% amylopectin by weight. Glycogen, the glucose store of animals, is a more branched version of amylopectin.

Section 1.4 – Amylose and Amylopectin

1. Amylose is a linear polymer made up of D-glucose units. This polysaccharide is one of the two components of starch, making up approximately 20-30% of the structure. The other component is amylopectin, which makes up 70-80% of the structure. Because of its tightly packed structure, amylose is more resistant to digestion than other starch molecules and is therefore an important form of resistant starch, which has been found to be an effective prebiotic. (Revedin, A. et al. (2010).)

Function

Amylose is important in plant energy storage. It is less readily digested than amylopectin; however, because it is more linear than amylopectin, it takes up less space. As a result, it is the preferred starch for storage in plants. It makes up about 30% of the stored starch in plants, though the specific percentage varies by species. The digestive enzyme α-amylase is responsible for the breakdown of the starch molecule into maltotriose and maltose, which can be used as sources of energy.

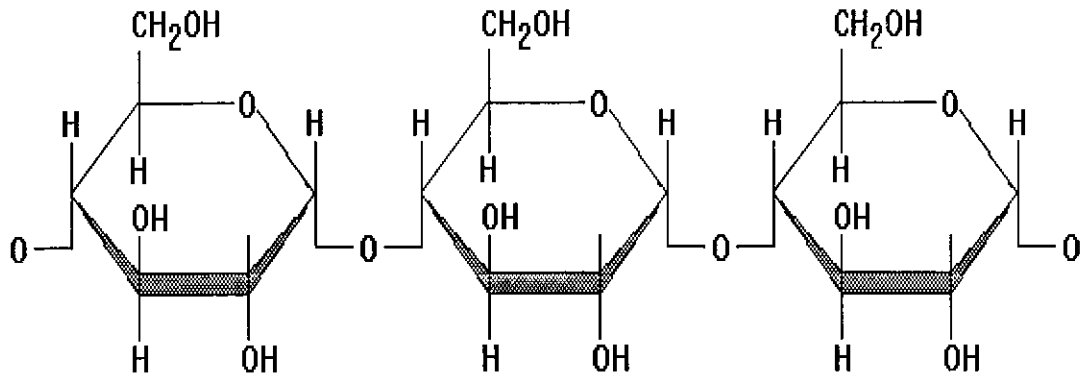


Fig 2: Amylose Structure

2. Amylopectin is a soluble polysaccharide and highly branched polymer of glucose found in plants. It is one of the two components of starch, the other being amylose.

   Glucose units are linked in a linear way with α (1→4) glycosidic bonds. Branching takes place with α(1→6) bonds occurring every 24 to 30 glucose units, resulting in a soluble molecule that can be quickly degraded as it has many end points for enzymes to attach onto.

   In contrast, amylose contains very few α(1→6) bonds, or even none at all. Its counterpart in animals is glycogen, which has the same composition and structure, but with more extensive

   Branching that occurs every 8 to 12 glucose units. (Revedin, A et al. (2010).)

Function

Amylose is important in plant energy storage. It is less readily digested than amylopectin; however Plants store starch within specialized organelles called amyloplasts. When energy is needed for cell work, the plant hydrolyzes the starch, releasing the glucose subunits. Humans and other animals that eat plant foods also use amylase, an enzyme that assists in breaking down amylopectin.

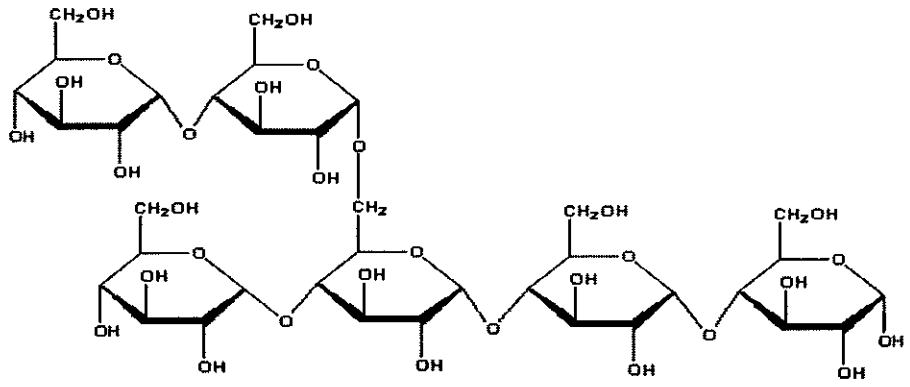


Fig 3: Amylopectin Structure

Starch Metabolism

Rice (*Oryza sativa* L.) is the staple food for more than half of the world's population. Each year many new rice varieties are being developed and released with high yield, and high resistance to biotic and a biotic stresses, but their eating quality remains to be improved.

Eating quality has now become the primary consideration of final rice-eating consumers and breeding programs. Cultivars with better eating quality are also required for food industry and rice seed producers. Although indica rice varieties are widely planted worldwide, consumers in Northeast Asia preferred japonica rice. Also, japonica rice is becoming more and more popular with worldwide consumers due to its moderate elasticity and stickiness [1]–[3].

Polished rice is mainly made up of starch, protein, lipid, and moisture. The protein content of polished rice in 22 japonica rice varieties ranged from 5.9 to 7.9% [3], and 6.0–13.6% in brown rice among 1,518 Chinese japonica varieties [4]. Starch comprises 76.7–78.4% in polished rice with 14% moisture content [5]. Eating quality of rice is thus mainly influenced by starch property. Moreover, eating quality of rice is a very complex trait, and the palatability from Toyo taste meter is significantly and positively correlated with palatability from sensory test [3]. Therefore, palatability is directly related to rice eating quality. In physicochemical level, apparent amylose content [6], gel consistency [7], gelatinization temperature [8] or alkali digestion value, pasting properties [9], chain length distribution of amylopectin [10], and protein content [11] are thought to be important parameters affecting eating quality. Rice eating quality is partially affected by environmental factors such as growing temperature and soil fertility, but mainly determined by genetic control [12].

In molecular biological aspects, sucrose derived from glucose and fructose is synthesized in cytoplasm, and then transported into cytosol in which sucrose is degraded by invertase. Using degraded products of sucrose, fructose and UDP-glucose, starch is synthesized by multiple subunits or isoforms of five classes of enzymes: ADP-glucose pyrophosphorylase (AGP), soluble starch synthase (SSS), granule-bound starch synthase (GBSS), starch branching enzyme (SBE), and starch debranching enzyme (DBE) [13]–[15] (Figure 1). Moreover, GBSS is not only responsible for amylose synthesis, but also involved in amylopectin synthesis, especially in forming the extra-long chains of amylopectin [16]. Among the five enzymes, SBE, GBSS, SSS, and DBE are contributing to the fine structure of amylopectin [17]. SBE in rice grain contains SBE1, SBE3 (QEIIa, or BEIIb), and SBE4 (QEIIb, or BEIIa) isoforms [18]–[20]. GBSS has two isoforms, GBSS1 in rice grain and GBSS2 in rice leaf [21]. SSS contains SSS1, SSS2A (SSSII-3), SSS2B (SSSII-2), SSS2C (SSSII-1), SSS3A (SSSIII-2), SSS3B (SSSIII-1), and SSS4A (SSSIV-1), SSS4B (SSSIV-2) [22]. DBE contains isoamylase (ISA) and pullulanase (PUL) in rice. ISA has at least three isoforms, ISA1, ISA2, and ISA3, and only one PUL [23].

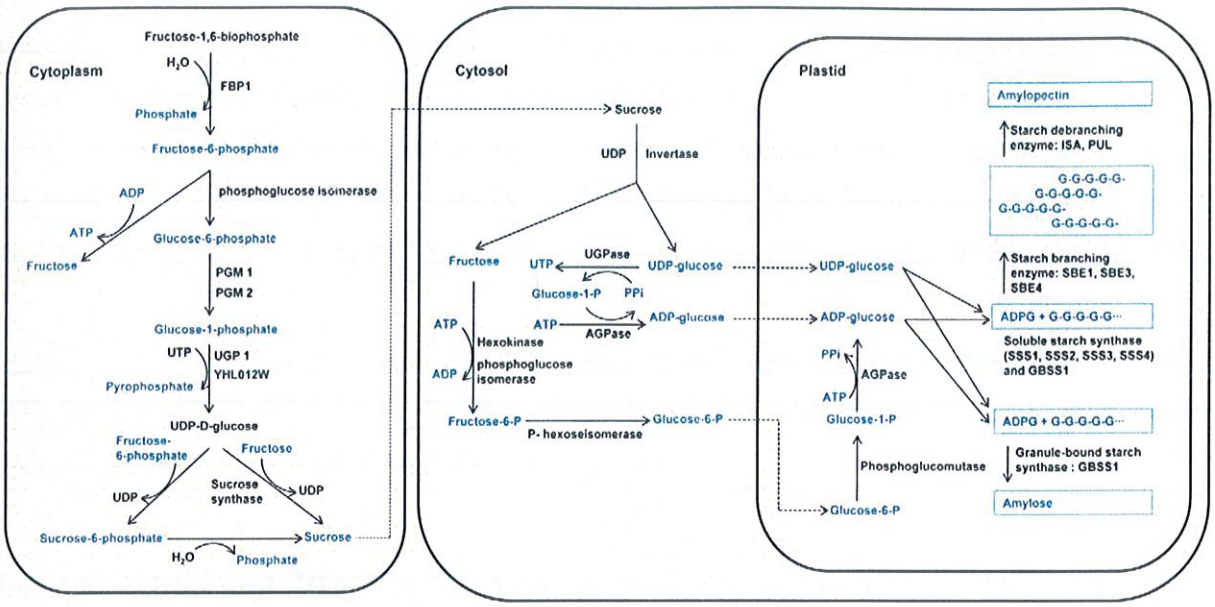


Fig 4.Main pathway of starch biosynthesis including sucrose synthesis, sucrose degradation and starch synthesis in rice.FBP1, fructose-1, 6-bisphophatase1; PGM

The enzymes and genes involved in fine structure of amylopectin play distinct roles, but jointly comprise of the genetic dissection of rice eating quality. Using 70 rice varieties, association analysis between 18 genes related to starch synthesis and amylose content, gel consistency, gelatinization temperature showed that Waxy (GBSS1) and ALK (SSS2A) are central in determining rice eating and cooking quality by affecting amylose content, gel consistency and gelatinization temperature [24]. Waxy functions as the sole major gene for both amylose content and gel consistency but as a minor gene affecting gelatinization temperature. ALK is the sole major gene controlling gelatinization temperature but as a minor gene affecting amylose content and gel consistency. The two genes affect two properties simultaneously: both ISA and SBE3 affect gel consistency and gelatinization temperature. Furthermore, several minor genes are specific for each property: SSS3A, AGPlar (AGP large subunit), PUL, andSSS1 for amylose content, AGPiso (AGP large subunit isoform) for gel consistency, andSSS4B for gelatinization temperature. The correlations among amylose content, gel consistency, and gelatinization temperature were caused by the joint action of these associated genes.

Amylose Content (AC)

Starch has two major components namely linear ά -polyglucan amylose and branched ά -polyglucan amylopectin. In general, stored starch in the higher plants is composed of 20-30% amylose and 70-80%

amylopectin. Rice varieties can be classified based on varied range of AC such as waxy (0-2%), very low (3-12%), low (13-20%), intermediate (21-25%) and high (≥26%) (Juliano et al., 1981). Rice grains with high AC (≥26%) cook dry, become less tender and hard upon cooling while rice grains with low AC (<20%) cook moist and become every sticky. That is why rice with intermediate AC (20-25%) is preferred in majority of the rice-growing/consuming regions of the world, except in few regions where only low-AC japonicas are preferred. Therefore, the efforts need to be intensified in developing perfect markers for differentiating each class of AC and marker-assisted introgression of desired allele, which will help the breeders in developing varieties with preferred AC range as per local and international consumer preference for rice cooking quality.

Section 1.5– GBSS and SBE

Granule bound starch synthase (GBSS)

GBSS enzyme comprises of two isoforms (GBSS-I, GBSS-II). The GBSS-I is encoded by waxy locus and is involved in the synthesis of long amylopectin (CL 85-180) chains (Takedaet al., 1987; Wang et al., 1995; Denyer et al., 1996; Cai et al., 1998; Fu and Xue, 2010) in higher proportion, resulting in the absence of very long chains in the waxy rice's (Hizukari etal., 1989). The expression of GBSS-I enzyme appears to be confined mostly to the storage tissues, while GBSS-II enzyme is encoded by another separate gene and is responsible for theamylose synthesis in the leaves (Fuzita and Taira, 1998; Vrinten and Nakamura, 2000; Cai et16al., 1998; Fu and Xue, 2010). The huge variation in rice germplasm for AC is determined by the activity level of the GBSS-I, which in turn depends on the 5 alleles (Wxa, Wxb, Wxin,Wxop and wx) (Sano, 1984; Mikami et al., 2008; Chen et al., 2008). A microsatellite (SSR with 8-20 CT repeats) in the un-translated region of exon 1 of the GBSS-I gene discriminates between different types of amylose (Ayres et al., 1997). A relationship between AC and number of CT repeats was established through several studies (Ayers et al., 1997; Bao et al.,2002; Bergman et al., 2001; Bligh et al., 1995; Olsen and Purugganan, 2002; Fitzgerald,2004) and been concluded as follow; (a) low amylose temperate Japonica rice's possess Wxaallele (18-19 CT) (b) medium amylose tropical japonica possess Wxin allele (11-10 CT repeats) and (c) high amylose Indica rice's carry Wxb allele (14-20 CT repeats).

Starch branching enzymes (SBE)

After elongation of the glucal chains by the SS enzyme, the another enzyme namely SBE with two isoforms i.e. SBE-I and SBE-II generates ά-(1-6) linkages by cleaving internal ά-(1-4) bonds and transferring the released reducing ends to C6 hydroxyls to form the branched structure of the amylopectin molecule. SBE-II proteins transfer shorter chains and show a higher affinity towards amylopectin as compared to the SBE-I, which show higher rates of branching with amylose (Guan and Preiss, 1993; Takeda et al., 1993; Tanaka et al., 2004;Tetlow et al., 2004; Tetlow, 2006; Nakamura et al., 2010) (Table 4). The termini (N- and C-) of these enzymes play important roles in determining the substrate preference, catalytic capacity and chain length transfer (Kuriki et al., 1997). In the monocots, the SBE-II gene has two closely related but distinct gene products i.e., SBE-IIa and SBE-IIb enzymes (Rahman etal, 2001). Down regulation or elimination of SBE-I gene activity showed minimal effects on the starch synthesis (Blauth et al., 2002; Satoh et al., 2003b; Flipse et al., 1996), while SBE17IIa gene showed clear/major role in the leaf starch synthesis and no effect on storage starch of endosperm (Blauth et al., 2001). The SBE-IIb enzyme has a distinct role in the transfer of short chains, which are then most probably extended by SS to form A and B1 chains of the rice amylopectin cluster structure (Nishi et al., 2001). Study on SBE-I mutants of rice have shown decreased level of the amylopectin chains with DP 12-20 and DP≥37; and increased level of amylopectin chains with DP 24-34 (Satoh et al., 2003). Hence, the role of the SBE-I enzyme is clear and important in the synthesis of B1, B2 and B3 amylopectin chains (Nakamura, 2002; Satoh et al., 2003). One more enzyme isoform (SBE III) has also been reported recently (Chen et al., 2004), which plays an important role in the synthesis of 1-6branching linkage.

GBSSI and SSIIa are major genes involved in many grain quality properties such as amylose content and gelatinization temperature. Highly significant associations were found between GBSSI and retro gradation and amylose content although this gene showed more significant relations with properties such as BDV, SB and FV. A number of authors have already reported the importance of this enzyme in determining the starch physiochemical properties in rice and other cereals. SNPs at the intron/exon 1 junction site, exon 6 and 10 in rice GBSSI (waxy gene) have the most significant impact on amylose content and by extension, starch quality8–10. This study confirms the 'T/G' SNP at the intron1/exon1 junction site has a major influence on a number of physiochemical properties. (SNP in starch biosynthesis genes associated with nutritional and functional

GBSS-1 and SBE

- GBSS-I(Granule Bound Starch Syntatase 1): important enzymes in Starch metabolism
- GBSS-I converts ADP-Glucose into Amylose during Starch Metabolism.
- SBE(Starch Branching Enzyme) – important enzymes in Starch Metabolism
- SBE converts ADP-Glucose into Amylopectin.

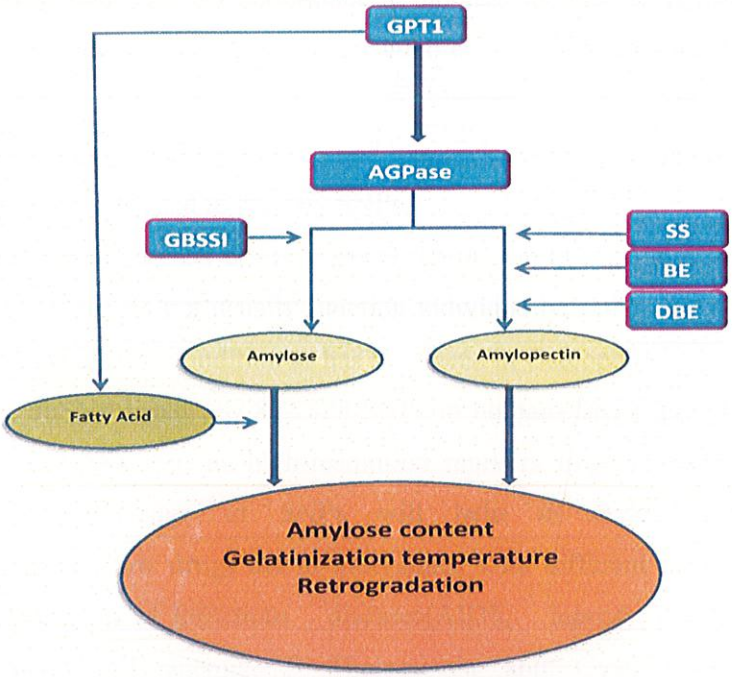


Fig5: Explains the importance of how GBSS 1 and SBE enzymes are directly responsible for Amylose and Amylopectin Content.

Section 1.6 – Single Nucleotide Polymorphism

A single-nucleotide polymorphism (SNP, pronounced *snip; plural snips*) is a DNAsequence variation occurring when a single nucleotide — A, T, CorG — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes in a human. For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. In this case we say that there are two *alleles*: C and T. Almost all common SNPs have only two alleles. The genomic distribution of SNPs is not homogenous, SNPs usually occur in non-coding regions more frequently than in coding regions or, in general, where natural selection is acting and fixating the allele of the SNP that constitutes the most favorable genetic

Use and importance

Variations in the DNA sequences of humans can affect how humans develop diseases and respond to pathogens, chemicals, drugs, vaccines, and other agents. SNPs are also critical for personalized medicine. [4] However, their greatest importance in biomedical research is for comparing regions of the genome between cohorts (such as with matched cohorts with and without a disease) in genome-wide association studies.

The study of SNPs is also important in crop and livestock breeding programs. See SNP genotyping for details on the various methods used to identify SNPs.

SNPs are usually biallelic and thus easily assayed. [5] A single SNP may cause a Mendelian disease. For complex diseases, SNPs do not usually function individually; rather, they work in coordination with other SNPs to manifest a disease condition as has been seen in Osteoporosis. [6]

As of 26 June 2012, dbSNP listed 187,852,828 SNPs in humans.SNPs have been used in genome-wide association studies (GWAS), e.g. as high-resolution markers in gene mapping related to diseases or normal traits. The knowledge of SNPs will help in understanding pharmacokinetics (PK) or pharmacodynamics, i.e. how drugs act in individuals with different genetic variants. A wide range of human diseases like cancer, infectious diseases(AIDS, leprosy, hepatitis, etc.) autoimmune, neuropsychiatric, Sickle–cell anemia, β Thalassemia and Cystic fibrosis might result from SNPs.[8][9][10] Diseases with different SNPs may become relevant pharmacogenomic targets for drug therapy.[11] Some SNPs are associated with the metabolism of different drugs.[12][13][14] SNPs without an observable impact on the phenotype are still useful as genetic markers in genome-wide association studies, because of their quantity and the stable inheritance over generations.[15]

Databases

As there are for genes, bioinformatics databases exist for SNPs. *dbSNP* is a SNP database from the National Center for Biotechnology Information (NCBI). *SNPedia* is a wiki-style database supporting personal genome annotation, interpretation and analysis. The *OMIM* database describes the association between polymorphisms and diseases (e.g., gives diseases in text form), the Human Gene Mutation Database provides gene mutations causing or associated with human inherited diseases and functional SNPs, and GWAS Central allows users to visually interrogate the actual summary-level association data in one or more genome-wide association studies. The International SNP Map working group mapped the sequence flanking each SNP by alignment to the genomic sequence of large-insert clones in Gene bank.

These Alignments were converted to chromosomal coordinates that is show in Table 1 [19] Another database is the International Hap Map Project, where researches are identifying Tag SNP to be able to determine the collection of haplotypes present in each subject.

SNP analysis

Analytical methods to discover novel SNPs and detect known SNPs include:

- DNA sequencing;[23]
- capillary electrophoresis;[24]
- mass spectrometry;[25]
- single-strand conformation polymorphism (SSCP);
- electrochemical analysis;
- denaturating HPLC and gel electrophoresis;
- restriction fragment length polymorphism;
- hybridization analysis;

Section 1.7 – Review of Literature

ASIAN cultivated rice (*Oryza sativa* L.) holds a unique position among domesticated crop species in that it is both a critical food staple and the first fully sequenced crop genome. Rice is consumed as a grain almost exclusively by humans, supplying 20% of daily calories for the world population (World Rice Statistics, http://www.irri.org; FAOSTAT, http://apps.fao.org). As a model organism with a fully sequenced genome, rice affords unique opportunities to use genomic approaches to study its domestication, adaptive diversity, and the history of crop improvement. (Genetic Structure and Diversity in *Oryza sativa* L. Amanda J. Garris,*,1 Thomas H. Tai,†,2 Jason Coburn,* Steve Kresovich,* and Susan McCouch*,3)

The literature pertinent to the present study has been reviewed under the following heads:

1. SNP in starch biosynthesis genes associated with nutritional and functional properties of riceFatty acids and TAG biosynthesis in plants
2. Comparison of Two Models to Predict Amylose Concentration in Rice Flours as Determined by Spectrophotometric Assay"

3. Effect of cooking on amylose content of rice
4. Disease resistance genes in plants

1. SNP in starch biosynthesis genes associated with nutritional and functional properties of rice

Fatty acids and TAG biosynthesis in plants

Starch is a major component of human diets. The relative contribution of variation in the genes of starch biosynthesis to the nutritional and functional properties of the rice was evaluated in a rice breeding population. Sequencing 18 genes involved in starch synthesis in a population of 233 rice breeding lines discovered 66 functional SNPs in exonic regions. Five genes, AGPS2b, Isoamylase1, SPHOL, SSIIb andSSIVb showed no polymorphism. Association analysis found 31 of the SNP were associated with differences in pasting and cooking quality properties of the rice lines. Two genes appear to be the major loci controlling traits under human selection in rice, GBSSI (waxy gene) and SSIIa. GBSSI influenced amylose content and retro gradation. Other genes contributing to retro gradation were GPT1, SSI, BEI and SSIIIa. SSIIa explained much of the variation in cooking characteristics. Other genes had relatively small effects. High-amylose rice cultivars usually have more resistant starch (RS) and lower estimated glycemic index (EGS); suggesting highly-retrograded cooked rice cultivars tend to a reduction of hydrolysis index (HI) and glycemic index (GI) 6. Conversely, starch of low-amylose rice's, which have higher HI, are more quickly hydrolyzed than intermediate and high-amylose rice (high HI)6,7. Characteristics of high amylose rice cultivars are normally determined by RVA (Rapid Visco Analysis) which are described by parameters such as peak viscosity (PKV), hot paste viscosity (HPV) and cool paste viscosity (CPV). Seven starch synthesis enzyme classes have been defined, including ADP-glucose pyrophosphorylase (AGPase), granule bound starch synthase (GBSS), starch synthase (SS), branching enzyme (BE), debranching enzyme (DBE), starch phosphorylase (PHO) and glucose 6-phosphate translocator (GPT). These genes/enzymes contribute directly or indirectly to the production of starch granules. The link between natural variation in particular starch synthesis genes and starch properties is well established in some cases. GBSSI (waxy gene) is primarily responsible for the synthesis of linear chains of glucose molecules found in amylose is the most well characterized cereal grain starch synthesis enzyme. A number of SNP in the rice waxy gene, at the intron/exon 1 junction site, exon 6 and exon 10, impact starch quality8–10 by effecting amylose content. The gene encoding starch synthase IIa (SSIIa), alk, is exclusively expressed in the rice endosperm and has been extensively studied in the context of its effect on cooking quality and starch texture11,12. Two SNPs within exon 8, [A/G] and [GC/TT] are significantly associated with rice alkali disintegration and eating quality and starch gelatinization temperature (GT) 13.More recently, Yan et al. (2010) analyzed the association of 17 starch synthesis

genes with RVA profile parameters in a collection of 118 glutinous rice accessions using 43 gene-specific molecular markers. They found 10 of 17 starch-related genes have an impact on rapid visco analyzer (RVA) profile parameters. The association analysis revealed pullulanase plays a dominant role in control of PKV, HPV, CPV, breakdown viscosity (BDV), peak time (PKT), and pasting temperature (PT) in glutinous rice. Nine other starch genes had a minor impact on only a few RVA profile parameters. However, RVA parameters such as starch paste viscosity and other starch quality traits may be controlled by a complex genetic system involving many starch-related genes14. Many induced mutations that have been (Ardashir Kharabian-Masouleh1, Daniel L. E. Waters1, Russell F. Reinke2, 3, Rachelle Ward4& Robert J. Henry5)

2. **Comparison of Two Models to Predict Amylose Concentration in Rice Flours as Determined by Spectrophotometric Assay".**

A spectrophotometric assay was used to test two models for predicting (<6%). The amylose-only method predicted amylose content to be 7.4% amylose in 16 solutions containing known concentrations of amylose and for waxy flour and 16.4-25.4% for no waxy varieties. Defatted floursamylopectin and in 10 rice flours of unknown composition. The simple were measured at 8.0% (waxy) and 18.4-29.6% (nonwaxy). Lower amylose linear regression model based on amylose-only standard curve over- content was measured by the simultaneous estimation method. Predictions predicted amylose in all solutions (relative bias of 7-329%). The model for the flours were 0% (waxy) and 6.6-14.2% (nonwaxy). Defatted flours used for simultaneous estimation of amylose and amylopectin was more had 0% (waxy) and 9.8-22.0% (nonwaxy). This method resulted in greater accurate with some slight over prediction or under prediction of amylose accuracy, although use of laboratory time and resources was not changed. Two methods were used to predict the concentration of amylose in 16 mixtures of known amylose and amylopectin concentration and in 10 flours of unknown composition. The weighted least squares regression based on an amylose-only standard curve over predicted amylose in all known mixtures (relative bias 7-329%). The simultaneous spectrophotometric method was a more accurate predictor of true amylose content, with relative biases being scattered about zero (-3 to 6%). This study showed that the most commonly used method to determine amylose content in rice flours is systematically biased toward over prediction. Researchers can improve accuracy by use of the simultaneous estimation method. Results are closer to the true concentration, especially at low amylose levels. Moreover, use of the simultaneous method involves no increase in use of laboratory resources. (P. S. LANDERS, 3 E. E. GBUR, 4 and R. N. SHARP3

3. Effect of cooking on amylose content of rice

In diabetes type 2, there is a deficiency of insulin which results in improper/ slow breakdown of food. This results in sugar level spikes immediately after a meal, which can be harmful. Hence, diabetics must have food stuffs which have a slow release rate so as to not cause any spikes. In order to find out the most suitable method of cooking rice for diabetic patients, we found out the amylose content of rice cooked by different methods and co-related it with Glycemic Index (GI). Glycemic Index is release of glucose in the blood by the breakdown of carbohydrates. Higher the glycemic index, faster breakdown of food and thereby more release of glucose in the blood, so immediate requirement of insulin, which can be toxic for diabetic patients. Glycemic index and amylose content are inversely proportional to each other. There is a wide variation in the amylose content of rice depending on the way it is cooked. In this paper, the effect of cooking on amylose content of rice is described using various experimental approaches. Various method of cooking involves traditional method, microwave and steam cooked method. The amylose content of the rice is then co-related with its glycemic index. (Ashish Jain, Sughosh M Rao, SarikaSethi, Abhinav Ramesh, SwapnilTiwari, Sanjeeb Kumar Mandal, Nishantkumar Singh, Ashish Singhal, NaisargModi, VibhavBansal and ChitraKalaichelvani)

4. An SNP Resource for Rice Genetics and Breeding Based on Subspecies Indica and Japonica Genome Alignments

Dense coverage of the rice genome with polymorphic DNA markers is an invaluable tool for DNA marker-assisted breeding, positional cloning, and a wide range of evolutionary studies. We have aligned drafts of two rice subspecies, indica and japonica, and analyzed levels and patterns of genetic diversity. After filtering multiple copy and low quality sequence, 408,898 candidate DNA polymorphisms SNPs/INDELs) were discerned between the two subspecies. These filters have the consequence that our data set includes only a subset of the available SNPs (in particular excluding large numbers of SNPs that may occur between repetitive DNA alleles) but increase the likelihood that this subset is useful: Direct sequencing suggests that 79.8% ± 7.5% of the in silico SNPs are real. The SNP sample in our database is not randomly distributed across the genome. In fact, 566 rice genomic regions had unusually high (328 contigs/48.6 Mb/13.6% of genome) or low (237 contigs/64.7 Mb/18.1% of genome) polymorphism rates. Many SNP-poor regions were substantially longer than most SNP-rich regions, covering up to 4 Mb, and possibly reflecting introgression between the respective gene pools that may have occurred hundreds of years ago. Although 46.2% ± 8.3% of the SNPs differentiate other pairs of japonica and indica genotypes, SNP rates in rice were not predictive of evolutionary rates for corresponding genes in another grass species, sorghum. The data set is freely available athttp://www.plantgenome.uga.edu/snp.

The discovery of large numbers of single nucleotide polymorphisms (SNPs) in genome-scale sequencing initiatives opens new doors into the study of the genome-wide distribution of diversity and its significance. SNP markers open the possibility of "variation maps" at 100-fold higher resolution than current collections of DNA polymorphisms. This potential is being realized in humans (Sachidanandam et al. 2001) and Arabidopsis (Jander et al. 2002; Schmid et al. 2003; Torjek et al. 2003), in which SNP marker density is measured on a kilobase scale, whereas current genetic maps in major crops only begin to break the megabase barrier. Owing to this high density and advanced genotyping technologies, it is possible to apply these markers to genome-scale linkage disequilibrium and association studies as well as provide a tool for applications such as DNA marker-assisted breeding.

The recent sequencing of representatives from two diverse "subspecies" of rice provides an early glimpse into the genomic structure of variation in a monocot plant. After Arabidopsis (a dicot), the rice (*Oryza sativa*) genome is the most completely sequenced plant genome. Fortuitously, rice genomes have been sequenced from two rice subspecies (indica and japonica), which are thought to have diverged more than 1 million years ago (Bennetzen 2000). The International Rice Genome Sequencing Project (IRGSP) has used a BAC-based strategy to sequence rice ssp. japonica cv. Nipponbare (Sasaki and Burr 2000), and assembled sequences have been published for rice Chromosomes 1 (Sasaki et al. 2002), 4 (Feng et al. 2002), and 10 (Rice Chromosome 10 Sequencing Consortium 2003). Recently, The Institute for Genomic Research (TIGR; http://www.tigr.org) released a tentative assembly of all 12 chromosomes. Syngenta has sequenced the same japonica cultivar using a shotgun approach (Goff et al. 2002). The Beijing Genomics Institute (BGI) has sequenced an indica subspecies using a shotgun approach (ssp. indica cv. 93-11; Yu et al. 2002). The availability of extensive sequence for each of the two rice subspecies offers a unique and rich comparative genomics opportunity.

In this study, we have aligned indica contigs to the japonica genome assembly for the purpose of providing an SNP resource useful for rice breeding and genetics. This provides about a 100-fold increase relative to prior SNP studies (Nasu et al 2002), in the sampling of loci and SNP alleles available to study genome-wide patterns of genetic and physical distributions of SNP variation in rice. We investigate the predictive value of levels of SNP variation in the sequenced strains to other rice genotypes, and to other monocots (e.g., Sorghum). This data set provides a generalized framework useful to perform high-density marker studies involving indica/japonica crosses, and also has application to many evolutionary and/or functional studies.(F. Alex Feltus,1 Jun Wan,1 Stefan R. Schulze,1 James C. Estill,1 Ning Jiang,2 and Andrew H. Paterson1,2,3)

Chr	Length[a]	Intervals[b]	Contigs	Mb	Intervals[b]	Contigs	Mb	Intervals[b]	Contigs	Mb	% Chr	Moran's I[c]
1	42.9	23	20	2.3	59	26	5.9	82	46	8.2	19.10%	0.372***
2	35.5	23	22	2.3	48	19	4.8	71	41	7.1	20.00%	0.370***
3	35	37	25	3.7	61	23	6.1	98	48	9.8	28.00%	0.499***
4	34.4	70	38	7	79	27	7.9	149	65	14.9	43.30%	0.474***
5	28.5	45	33	4.5	79	16	7.9	124	49	12.4	43.50%	0.489***
6	29.9	68	33	6.8	58	18	5.8	126	51	12.6	42.10%	0.554***

Table 2: Nonrandom Distribution of *Indica-Japonica* Polymorphic Regions

A Megabase. B 100 kb. C Measurement of global spatial autocorrelation; ***indicates significance ($P < 0.001$).

Section 1.7: OBJECTIVES

a. Development of Universal Primers for GBSS-I and SBE.

b. Putative SNP identification for GBSS-I and SBE.

c. Association Analysis for finding Rule discovery.

CHAPTER-2

Section 2.1 – Material and Methods

20 subspecies of Indica Variety Rice (Table 1) were procured from NBPGR Shimla and NBPGR New Delhi

S/No	Variety Name
1	Navri Nut (white)
3	JatooDhan
S1	Basmati
5	Ram Jawain
6	Kala Dhan
7	Navri Nut Lal
8	PhulPatash
9	Rangoli
10	ZiriDhan
11	LalDhan
12	ChuhetuDhan
14	Rhodu White Dhan
15	KardDhan
S2	S2 Sugandh seed
S3	S3 Basmati Seed
S4	S4 Duplicate Basmati Seed
S5	S5 RR Seed

Fig 6:Rice Genotypes

- Breaking Seed Dormancy: Gibberlin (200ppm) sample was used to break the seed dormancy of those Rice Cultivars which did not show growth.

- Primer Designing: Primer Designing was done by using the software Primer 3 for developing Overlapping Primers. The Details of All the primers developed for GBSS 1 and SBE is discussed in Primer Sequence Table
- **Genomic DNA isolation:**
- CTAB buffer 100ml
    - 2.0 g CTAB (Hexadecyltrimethyl-ammonium bromide)
    - 10.0 ml 1 M Tris pH 8.0
    - 4.0 ml 0.5 M EDTA pH 8.0 (EthylenediaminetetraAcetic acid Di-sodium salt)
    - 28.0 ml 5 M NaCl
    - 40.0 ml H2O
    - 1 g PVP 40 (polyvinyl pyrrolidone (vinylpyrrolidinehomopolymer) Mw 40,000)
    - Adjust all to pH 5.0 with HCL and make up to 100 ml with H2O.

- 1 M Tris pH 8.0
    - Dissolve 121.1 g of Tris base in 800 ml of H2O. Adjust pH to 8.0 by adding 42 ml of concentrated HCL.
    - Allow the solution to cool to room temperature before making the final adjustments to the pH. Adjust the volume to 1 L with H2O. Sterilize using an autoclave.
- 5x TBE buffer
    - 54 g Tris base
    - 27.5 g boric acid
    - 20 ml of 0.5M EDTA (pH 8.0)
    - Make up to 1L with water.
    - To make a 0.5x working solution, do a 1:10 dilution of the concentrated stock.
- 1% Agarose gel
    - 1 g Agarose dissolved in 100 ml TBE
- Procedure
    - Grind 200 mg of plant tissue to a fine paste in approximately 500 µl of CTAB buffer.
    - Transfer CTAB/plant extract mixture to a microfuge tube.
    - Incubate the CTAB/plant extract mixture for about 15 min at 55o C in a recirculatingwater bath.
    - After incubation, spin the CTAB/plant extract mixture at 12000 g for 5 min to spin

down cell debris. Transfer the supernatant to clean microfuge tubes.

- To each tube add 250 µl of Chloroform: Iso Amyl Alcohol (24:1) and mix the solution by inversion. After mixing, spin the tubes at 13000 rpm for 1 min.
- Transfer the upper aqueous phase only (contains the DNA) to a clean microfuge tube.
- To each tube add 50 µl of 7.5 M Ammonium Acetate followed by 500 µl of ice cold absolute ethanol.
- Invert the tubes slowly several times to precipitate the DNA. Generally the DNA can be seen to precipitate out of solution. Alternatively the tubes can be placed for 1 hr at-20 o C after the addition of ethanol to precipitate the DNA.
- Following precipitation, the DNA can be pipetted off by slowly rotating/spinning a tip in the cold solution. The precipitated DNA sticks to the pipette and is visible as a clear thick precipitate.
- To wash the DNA, transfer the precipitate into a microfuge tube containing 500 µl of ice cold 70 % ethanol and slowly invert the tube. Repeat. ((Alternatively the precipitate can be isolated by spinning the tube at 13000 rpm for a minute to form a pellet.
- Remove the supernatant and wash the DNA pellet by adding two changes of ice cold 70 % ethanol)).
- After the wash, spin the DNA into a pellet by centrifuging at 13000 rpm for 1 min.
- Remove all the supernatant and allow the DNA pellet to dry (approximately 15 min).Do not allows the DNA to over dry or it will be hard to re-dissolve. Resuspend the DNA in sterile DNase free water (approximately 50-400 µl H2O; the amount of water needed to dissolve the DNA can vary, depending on how much is isolated).

- **Gel Extraction Protocol**

1. Excise the DNA fragment from the agarose gel with a clean, sharp scalpel or blade. (Minimize the size of the gel slice by removing extra agarose)
2. Weigh the gel slice in a colorless tube. Add 3 volumes of buffer MG to1 volume of gel. (100mg ~ 100µl). For example, add 300µl of buffer MG to each 100mg of gel.
3. Incubate at 50 °C in a water bath for 10 minutes (or until the gel slice has completely dissolved). To help dissolve the gel, mix by vortexing the tube every 2-3 minutes, during incubation.
4. After the gel slice has dissolved completely, add 1 gel volume of Isopropanol to the sample and mix by inverting the tube. For example, if the agarose gel weight is 100 mg, add 100µl isopropanol.

5. Place the spin column in the collection tube in a suitable rack. To bind DNA, load the sample to the spin column and centrifuge for 1 minute at >10,000 rpm. The maximum volume of the column reservoir is 800µl. For sample volumes of more than 800µl, simply load, balance and spin again.
6. Discard the flow through, and using the same collection tube, wash the spin column with 500µl of buffer MG by centrifuging for 1 minute at >10,000 rpm.
7. Discard the flow through, and using the same collection tube, wash the spin column with 750µl of buffer GW by centrifuging for 1 minute at >10,000 rpm.
8. Discard the flow through, place the spin column in the same collection tube and centrifuge for 1-2 minutes at >10,000 rpm.
9. Place the spin column in a fresh 1.5ml microfuge tube (Not provided). Elute the DNA by dispensing 30µl of elution buffer GE directly on to the center of the column membrane, allow it to stand for 1 minute and centrifuge for 1 minute at >10,000 rpm. Note: The elution buffer should be dispensed on to the center of the column membrane for maximum recovery. The average eluate volume is 28µl from 30µl elution buffer. If the eluate volume is lower, then centrifuge for one minute extra at >10,000 rpm.

- **Amylose Quantification Protocol:**

The following reagents are used for estimation of amylose content and for preparation of standard:

- 95% Ethanol: Prepared from 100% Ethanol
- 1N NaOH, Iodine - Potassium iodide solution
- Standard amylose: Obtained from HIMEDIA
- 1N Acetic acid: From NICE Chemicals
- Glassware: Borosil
- Water bath: High Precision water bath from Acmas was used
- Spectrophotometer: From Amersham Biosciences to measure OD
- Cuvette: Quartz Cuvette to measure OD
- Software: Handy Graph Software to draw graph
- 1N NaOH solution: Dissolve 40g of NaOH in 1000ml distilled water
- 1N Acetic acid solution: Dilute 57.5 ml glacial acetic acid to 1000ml using distilled water

- Iodine - Potassium iodide solution: Dissolve 0.26 g of Iodine in 10 ml of Potassium iodide solution containing 2.6 g of KI

Standard Amylose Solution: Take 40mg of pure potato starch (amylose) in a 100 ml volumetric flask and add 1 ml of 95% ethanol and 9.0 ml of 1N NaOH. Shake well and boil over water bath for 10 minutes and make up the solution to 100 ml using distilled water.

Method: We weighed 100 mg well powdered milled rice into 100 ml volumetric flask and to it 1 ml 95% ethanol and 9 ml 1 N NaOH was added. The sample was heated for 10 minutes in boiling water bath, cooled and the volume was made up to 100 ml. 5 ml was pipetted from the 100 ml into another 100 ml volumetric flask. To it 1 ml I N acetic acid and then 2 ml iodide solution were added and the volume was made up to 100 ml. The mixture was stirred and allowed to stand for 20 minutes and the per cent Transmittance at 620 nm was determined using a colorimeter. A series of standard starch solution containing 0, 20, 40, 60, 80 and 100% amylose was prepared as in the steps 1 to 5. The transmittance of the standards was read at 620nm and a standard graph was plotted. Amylose content of the sample was determined in reference to the standard curve and expressed on percent basis.

- **Primer Designing :**

  Software Used-Primer 3

  Overlapping primers were designed for GBSS 1, SBE1, SBE 3, SBE 4 genes.

PRIMER	SEQUENCE	PRODUCT	Avg
GBBS 1- 1F	GGTTGGAAGCATCACGAGTT	1863	57.85
GBBS 1-1R	GGGCTGGAGAAATCAACAAG		
GBBS 1-2F	CCCCTCTCTCACCATTCCTT	1034	59.85
GBBS 1-2R	AGTTTCTTGGGTGGCTAGGG		
GBBS 1-3F	GCATGAACGTCGTGTTCGTC	1100	59.85
GBBS 1-3R	TCTTCAGGTAGCTCGCCAGT		
GBBS 1-4F	GAGAAGTATCCGGGCAAGGT	872	58.85
GBBS 1-4R	CACATGTTTGACCGTTCGTC		
GBBS 1-5F	TGGGTTCGCTTCTCTTCTCT	623	57.85
GBBS 1-5R	TTTCCAGCCCAACACCTTAC		

Table 3: GBSS-1 Primer Sequence

PRIMER	SEQUENCE	PRODUCT	Tm
Sbe1 1F	AAGCTGTGGAAATGGGAGTC	863	59.14
Sbe1 1R	CACATTGTCTGCAAATTCTCTG		58.43
Sbe1 2F	TGAGTGGTGAAGAGCCAGAA	855	59.55
Sbe1 2R	TTGTCACCAACAATGGACTGA		59.99
Sbe1 3F	ACCGCAAATGGTCAATGAT	1101	60.38
Sbe1 3R	TTCTTTCTTTAAGGCGAACTATACA		57.93
Sbe3 1F	CCTCCTCGCTTTGGCTCT	841	60.63
Sbe3 1R	GGAGATGAGCCATCAGCATT		60.19
Sbe3 2F	TGTTTGGGAGATTTTTCTGC	865	57.79
Sbe3 2R	TGGCCTCAGGATAAAGTCCA		60.59
Sbe3 3F	CAAGTAGCATTTACGGGGAAC	1080	58.64
Sbe3 3R	TTTTCTTGCTCCTTCGCACT		60.13
Sbe4 1F	TTCTTCTTCTTCGACTTGACTCC	863	59.16
Sbe4 1R	GCATTTGGGTTCCAATTGTT		59.67
Sbe4 2F	CTACCGAGAATGGGCACCT	859	60.08
Sbe4 2R	ATGCCACCTGTAAACCATGA		58.85
Sbe4 3F	GATGGGTTTCGATTTGATGG	863	60.13
Sbe4 3R	CCAGTGGAAGTTGAACACGA		59.72
Sbe4 4F	GGTGGCATTTACTGGCAACT	1055	60.00
Sbe4 4R	TGAGTTGTTCCCCCAATGAT		60.17

Table 4: Primer for Starch Branching Enzyme

- **PCR Amplification**: PCR amplification was carried out for target gene(s) using the given parameters

S.No	PCR Chemicals	50ul Sample
1	Autoclave Water	36.5ul
2	Template	3ul
3	Primer(F)	2ul
4	Primer(R)	2ul
5	dNTPs	1ul
6	Taq Polymerase	0.5ul

7	Buffer	5ul

Table5: PCR parameters

- **PCR Programme :**
  - Conditions
    - Initial Denaturation Temp - 94°C. Time: 4 min on initial cycle;
    - Denaturation 30 seconds
    - Annealing -Temp: 5°C below Tm of primers; no lower than 40°C. Time: 30-45 seconds. This is the step where you would use a gradient.
    - Extension - Temp: 72°C. Time: 7 min/kb of expected product; 5-10 min on last cycle.
    - Number of Cycles - 35 cycles

- 50ul Sample of Amplified Gene - Sequencing
- Bio-Informatics Tools :
  - Primer Designing- Primer3
  - SNP finding – Snapster
  - Association Mining – Samiam
  - Amylose quantification – Regression Analyzer

CHAPTER-3

Section 3.1 – Results and Discussions

1. Growth of 20 varieties of Rice Seeds: Seeds of 20 different Genotypes of Rice were germinated at an optimum temperature of 37°C.
2. Genomic DNA Isolation

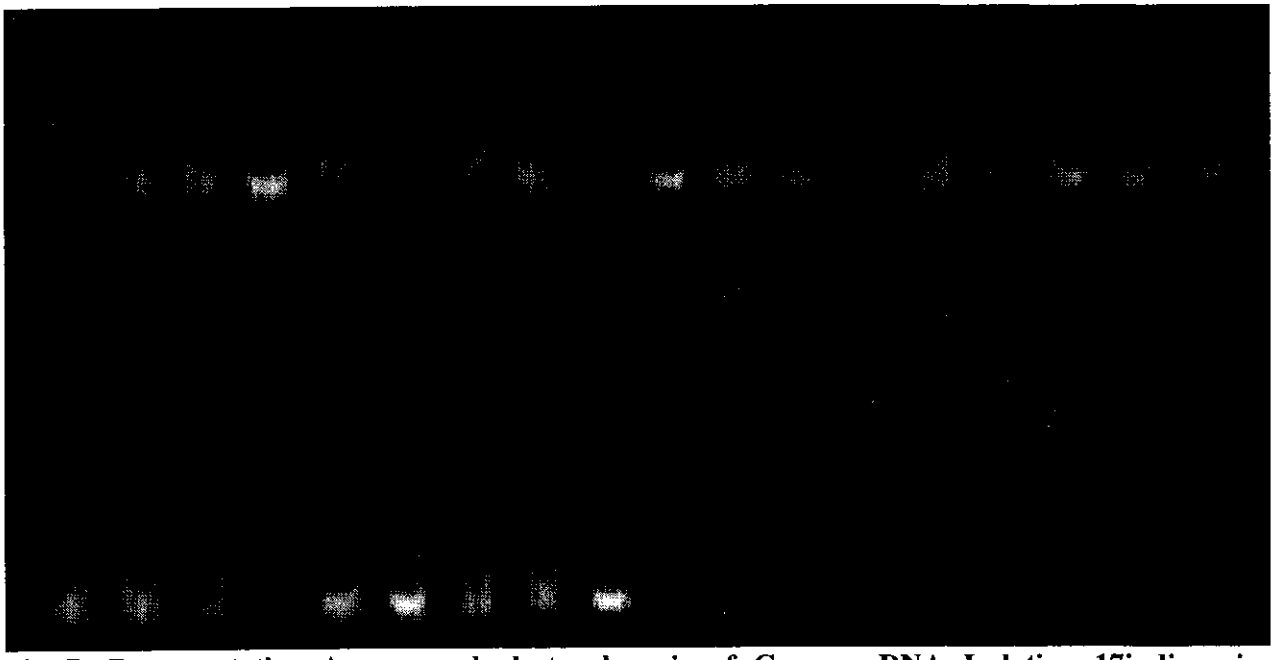


Fig 7: Representative Agarose gel electrophoresis of Genome DNA Isolation 17indica rice cultivars and prominent bands were observed.

3. **PCR Amplification :**

PCR Amplification for GBSS1 and Temperature Optimization

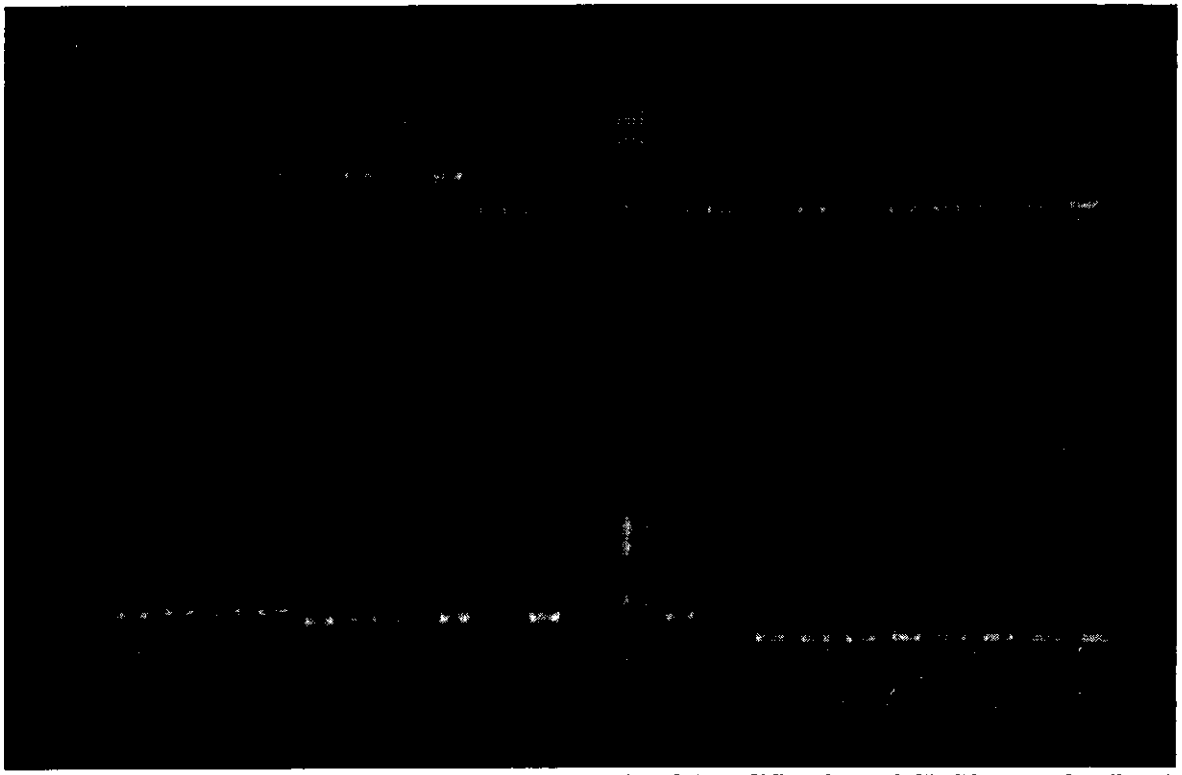


Fig 8: Representative Agarose gel electrophoresis of Amplification of GBSS gene for 5 pair primer set on 8 indica variety.

Quantification of Amylose:

Standard	A1 (Abs 620nm)	A2 (Abs 620nm)	A3 (Abs 620nm)	Avg(Absorbance)
Blank	0	0	0	0
1	0.0245	0.0265	0.0293	0.026766667
2	0.05	0.046	0.0499	0.048633333
3	0.0776	0.0691	0.0693	0.072
4	0.0989	0.0989	0.0996	0.099133333
5	0.1239	0.1288	0.1176	0.123433333

Table 6: This table is used in making the graph shown below. The procedure followed to obtain these readings is described in the Materials and Methods section. A1, A2, and A3 are the triplicate values.

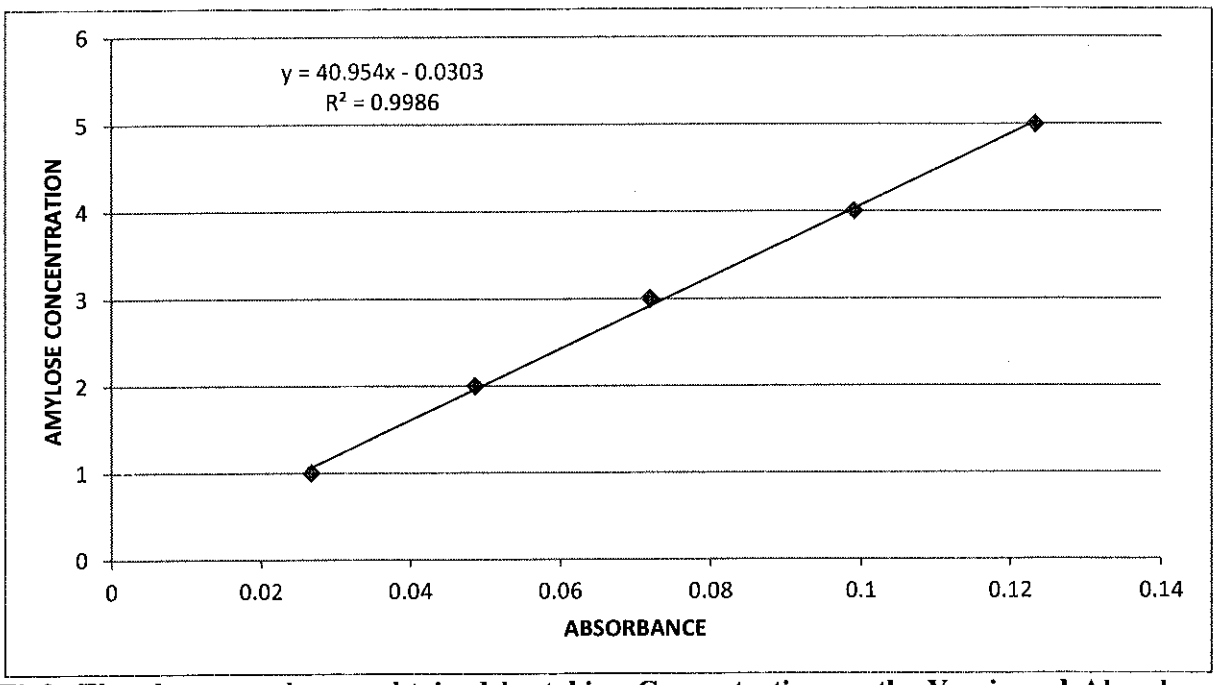


Fig9: The above graphs are obtained by taking Concentration on the Y-axis and Absorbance at 620nm on the X-axis for each of the set of values.

- Amylose Quantification on 20 Indica Varieties.

Sample	Absorbance (Abs 620nm)	Value of Y (mg/ml)	Characterization
S5	0.5479	22.406505	Intermediate
s2	0.216	8.8152	Very Low
S3	0.1671	6.812745	Very Low
S4	0.2696	11.01012	Very Low
7	0.189	7.70955	Very Low
1	0.2427	9.908565	Very Low
2	0.2357	9.621915	Very Low

3	0.3649	14.912655	Low
4	0.2243	9.155085	Very Low
5	0.3559	14.544105	Low
11	0.3886	15.88317	Low
12	0.322	13.1559	Low
13	0.346	14.1387	Low
14	0.2897	11.833215	Very Low
15	0.3756	15.35082	Low
6	0.2546	10.39587	Very Low
8	0.3444	14.07318	Low
9	0.2336	9.53592	Very Low
10	0.2419	9.875805	Very Low
16	0.3384	13.82748	Low

Table 7: Table lists the amylose content of 20 varieties of rice. The values are obtained from extrapolating from the standard amylose curve and based on which characterization is done.

Discussion:

If	Then
Amylose content is high	Low Glycemic Index and the rice grains will show high volume expansion (not necessarily elongation) and a high degree of flakiness. The rice grains cook dry, are less tender, and become hard upon cooling.
Amylose content is low	High glycemic Index and the rice grains will cook moist and sticky

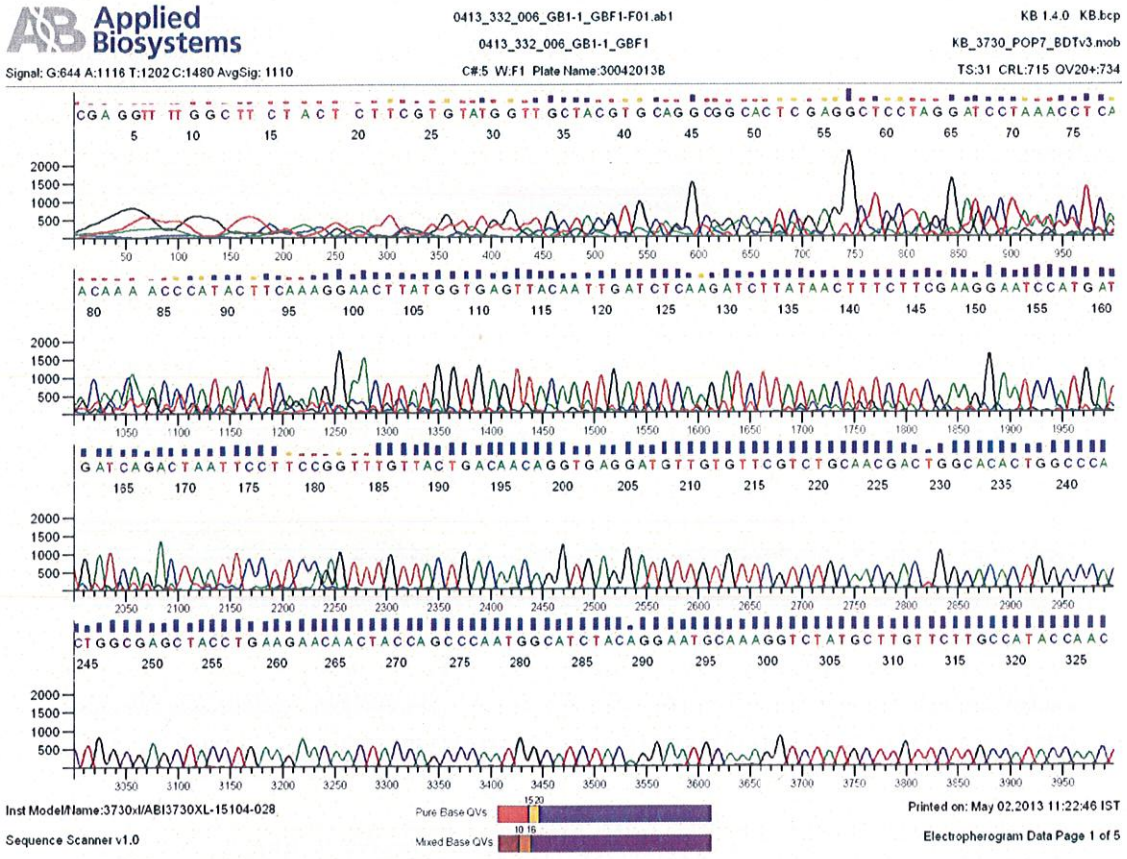
Table 8: Co-relation between amylose content and Glycemic index of rice

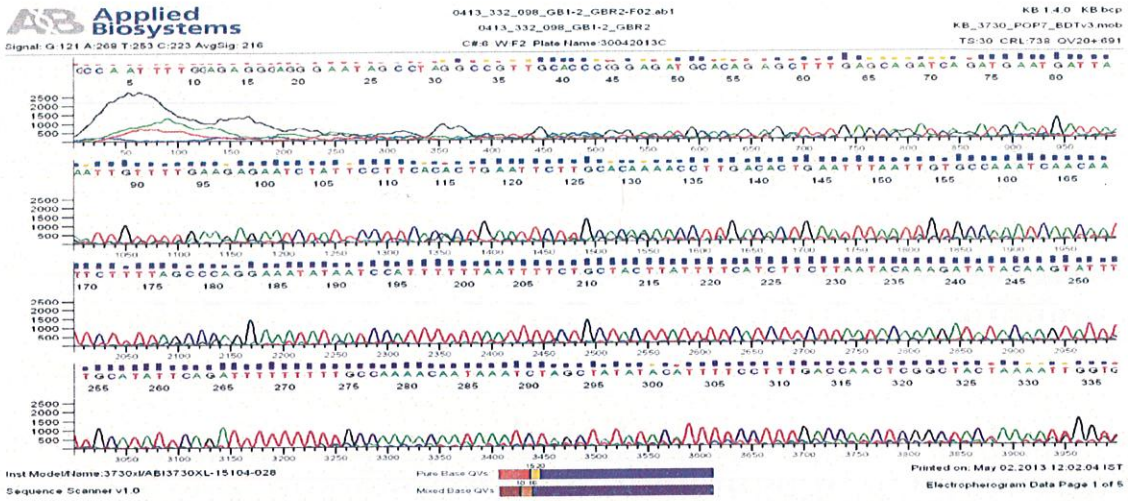
It is seen that amylose content is inversely related to the Glycemic index.

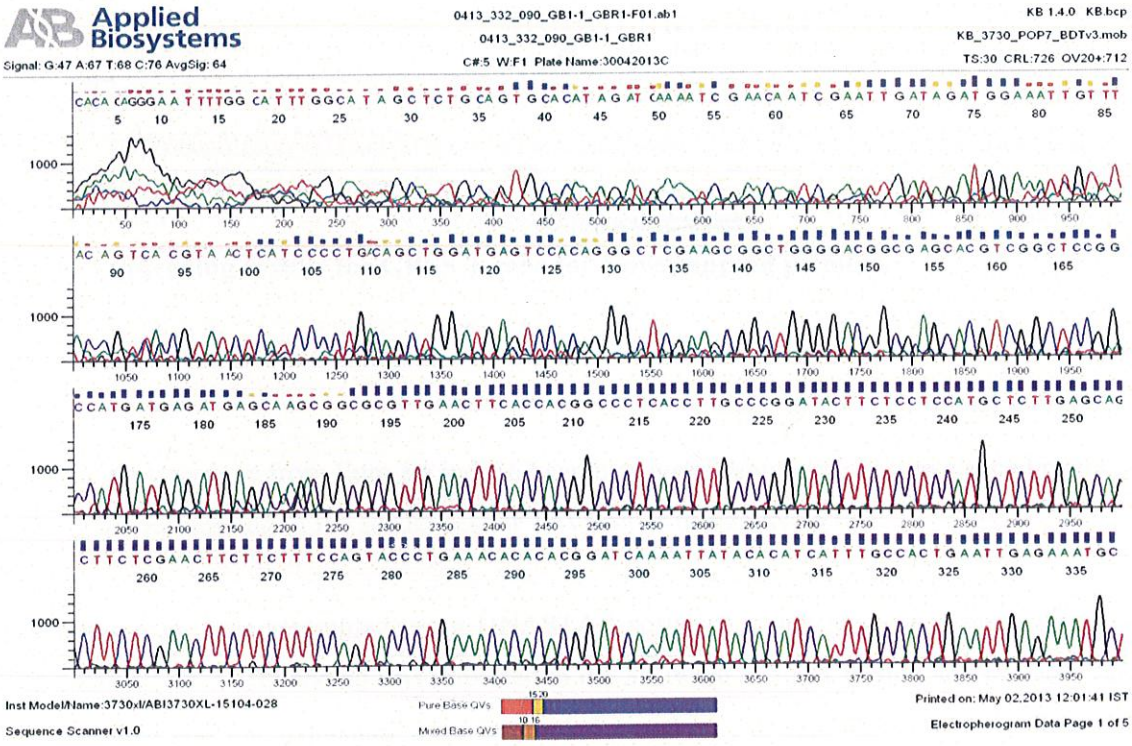
1. **Sequencing of GBSS-1 Genes** :

- Amplification of GBSS-1 gene was done on 17 varieties.
- Sequencing done from Euro fin India.
- 86 samples sent out of it 24 failed and 62 samples passed.

1. **Screenshots Of Sequenced Gene**




Applied Biosystems
0413_332_098_GB1-2_GBR2-F02.ab1
0413_332_098_GB1-2_GBR2
C#:6 W:F2 Plate Name:30042013C
KB 1.4.0 KB.bcp
KB_3730_POP7_BDTv3.mob
TS:30 CRL:738 QV20+:691
Signal: G:121 A:269 T:253 C:223 AvgSig: 216
Inst Model/Name:3730xl/ABI3730XL-15104-028
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: May 02,2013 12:02:04 IST
Electropherogram Data Page 1 of 5


Applied Biosystems
0413_332_090_GB1-1_GBR1-F01.ab1
0413_332_090_GB1-1_GBR1
C#:5 W:F1 Plate Name:30042013C
KB 1.4.0 KB.bcp
KB_3730_POP7_BDTv3.mob
TS:30 CRL:726 QV20+:712
Signal: G:47 A:67 T:68 C:76 AvgSig: 64
Inst Model/Name:3730xl/ABI3730XL-15104-028
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: May 02,2013 12:01:41 IST
Electropherogram Data Page 1 of 5

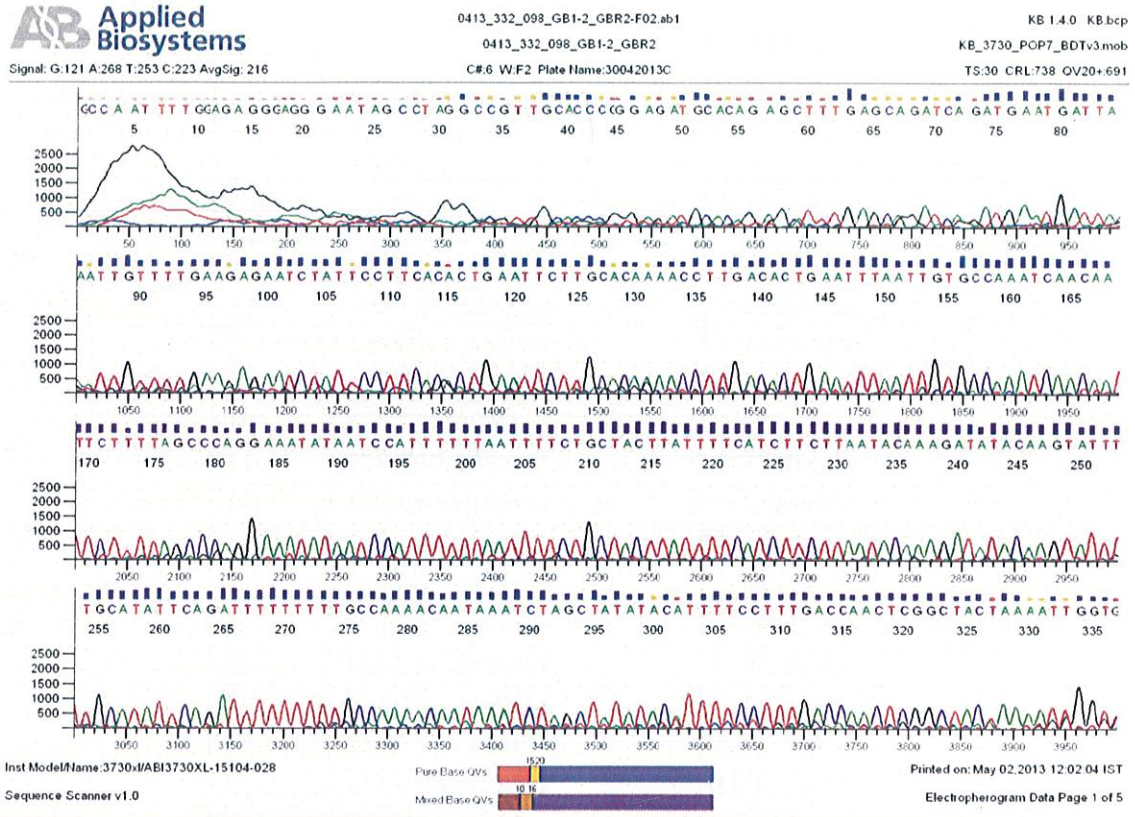


Fig10: Sequencing results for GBSS-1 gene for 5 overlapping primers.

Sequencing was done from Euro fin india.62 samples were sequenced. The sequenced results were further utilized for assembling of the sequences for 17 varieties using DNASTAR

The Sequences were assembled using DNASTAR software which has a tool Megalign. The file which was obtained after Sequencing was imported to the Software in *.ab1 format and for each variety, the end was trimmed and the sequences were assembled based on Forward and Reverse primer sequences and similarly assembled sequences for 16 more varieties were obtained. In order to validate the sequences, Blast Search was done with the *Oryza Sativa* Database and these results for obtained with following query coverage.

Rice	Query Coverage	Gaps
1	735/741(99%)	0/741(0%)
3	705/708(99%)	1/708(0%)
5	852/862(99%)	9/862(1%)
6	786/797(99%)	6/797(0%)
7	1006/1009(99%)	0/1009(0%)
8	1293/1295(99%)	2/1295(0%)
9	1374/1383(99%)	4/1383(0%)
10	550/552(99%)	2/552(0%)
11	729/730(99%)	0/730(0%)
12	1290/1296(99%)	2/1296(0%)
14	923/928(99%)	5/928(0%)
15	860/863(99%)	2/863(0%)
s1	1068/1079(99%)	1/1079(0%)
s2	1217/1218(99%)	0/1218(0%)
s3	1124/1128(99%)	3/1128(0%)
s4	1218/1218(100%)	0/1218(0%)
s5	1186/1208(98%)	18/1208(1%)

Table 11: Blast Results for Assembled Sequence

- **SNP Identification**:

  SNP identification was done by using SNIPlay web based tool in which the sequences were compared with *Oryza sativa* SNP database and the following results were obtained.

Pipeline for SNP Analysis

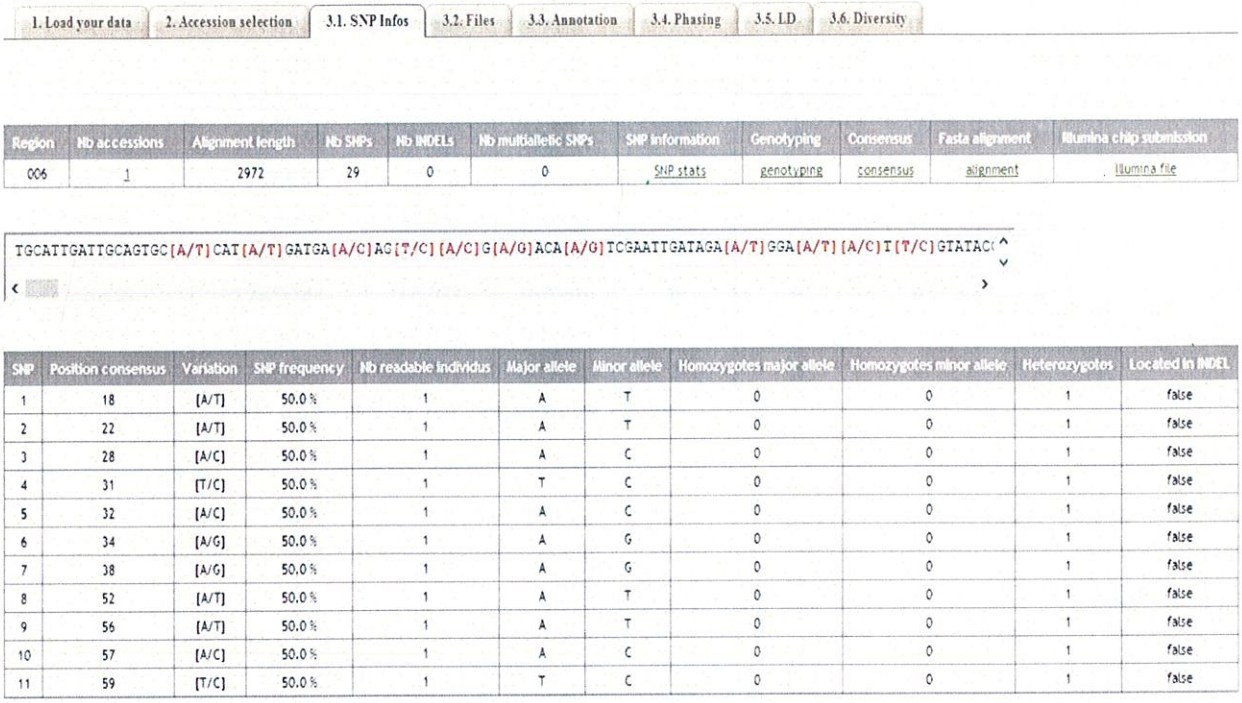


Fig 11: Screenshot for SNP Analysis using SNPiPlay

SNP	Chromosome	Position on chromosome	Variation	Feature	Position in codons	Codons	Amino acid	Effect		
1	chr06	1768581	[A/T]	intron	#	#	#	#		
2	chr06	1768585	[A/T]	intron	#	#	#	#		
3	chr06	1768591	[A/C]	intron	#	#	#	#		
4	chr06	1768593	[T/C]	intron	#	#	#	#		
5	chr06	1768594	[A/C]	intron	#	#	#	#		
6	chr06	1768596	[A/G]	intron	#	#	#	#		
7	chr06	1768600	[A/G]	intron	#	#	#	#		
8	chr06	1768614	[A/T]	exon						
9	chr06	1768618	[A/T]	exon	1	Tcc/Acc	S/T	NON_SYNONYMOUS_CODING		
10	chr06	1768619	[A/C]	exon	2	tCc/tAc	S/Y	NON_SYNONYMOUS_CODING		
11	chr06	1768621	[T/C]	exon						
12	chr06	1768631	[T/C]	exon	2	cTc/cCc	L/P	NON_SYNONYMOUS_CODING		
13	chr06	1768660	[C/G]	exon	1	Ggt/Cgt	G/R	NON_SYNONYMOUS_CODING		
14	chr06	1768725	[A/C]	intron	#	#	#	#		
15	chr06	1768727	[A/C]	intron	#	#	#	#		
16	chr06	1768867	[T/C]	exon	3	gtC/gtT	V/V	SYNONYMOUS_CODING		
17	chr06	1768883	[C/G]	exon	1	Gag/Cag	E/Q	NON_SYNONYMOUS_CODING		

18	chr06	1768895	[A/G]	exon	1	Agg/Ggg	R/G	NON_SYNONYMOUS_CODING		
19	chr06	1768899	[T/C]	exon						
20	chr06	1768912	[A/G]	exon	3	caG/caA	Q/Q	SYNONYMOUS_CODING		
21	chr06	1768918	[T/C]	exon	3	ctC/ctT	L/L	SYNONYMOUS_CODING		
22	chr06	1768921	[T/C]	exon	3	tcC/tcT	S/S	SYNONYMOUS_CODING		
23	chr06	1768968	[T/C]	intron	#	#	#	#		
24	chr06	1768982	[T/C]	intron	#	#	#	#		
25	chr06	1765008	[T/G]	intron	#	#	#	#		
26	chr06	1765013	[T/G]	intron	#	#	#	#		
27	chr06	1765041	[A/G]	intron	#	#	#	#		
28	chr06	1765042	[A/G]	intron	#	#	#	#		
29	chr06	1765045	[A/G]	intron	#	#	#	#		

Table 9: SNP annotation for sequenced Gene

SNP	Position consensus	Variation	SNP frequency	Nb readable individus	Major allele	Minor allele	Heterozygotes	SNP in INDEL
1	18	[A/T]	50.00%	1	A	T	1	FALSE
2	22	[A/T]	50.00%	1	A	T	1	FALSE
3	28	[A/C]	50.00%	1	A	C	1	FALSE
4	31	[T/C]	50.00%	1	T	C	1	FALSE
5	32	[A/C]	50.00%	1	A	C	1	FALSE
6	34	[A/G]	50.00%	1	A	G	1	FALSE
7	38	[A/G]	50.00%	1	A	G	1	FALSE
8	52	[A/T]	50.00%	1	A	T	1	FALSE
9	56	[A/T]	50.00%	1	A	T	1	FALSE
10	57	[A/C]	50.00%	1	A	C	1	FALSE
11	59	[T/C]	50.00%	1	T	C	1	FALSE
12	69	[T/C]	50.00%	1	T	C	1	FALSE
13	98	[C/G]	50.00%	1	C	G	1	FALSE
14	163	[A/C]	50.00%	1	A	C	1	FALSE
15	165	[A/C]	50.00%	1	A	C	1	FALSE
16	305	[T/C]	50.00%	1	T	C	1	FALSE
17	321	[C/G]	50.00%	1	C	G	1	FALSE
18	333	[A/G]	50.00%	1	A	G	1	FALSE
19	337	[T/C]	50.00%	1	T	C	1	FALSE
20	350	[A/G]	50.00%	1	A	G	1	FALSE
21	356	[T/C]	50.00%	1	T	C	1	FALSE
22	359	[T/C]	50.00%	1	T	C	1	FALSE
23	406	[T/C]	50.00%	1	T	C	1	FALSE
24	420	[T/C]	50.00%	1	T	C	1	FALSE
25	946	[T/G]	50.00%	1	T	G	1	FALSE

26	951	[T/G]	50.00%	1	T	G	1	FALSE
27	979	[A/G]	50.00%	1	A	G	1	FALSE
28	980	[A/G]	50.00%	1	A	G	1	FALSE
29	983	[A/G]	50.00%	1	A	G	1	FALSE

Table 10: SNP statistics for sequenced Gene

Amplicon	Haplotype name	Haplotype sequence	Size	Frequency
6	haplo1	TTCTCGGTTCTTGCCTGGTGTTTTTTGGG	1	50.00%
6	haplo2	AAACAAAAAACCCAACCACACCCCGGAAA	1	50.00%

Table 11 : Haplotype statistics for sequenced Gene

Discussion:

Insilico SNP Analysis was done using SNPiPlay webbased tool. Sequencing GBSS 1 gene involved in starch synthesis in a population of 17 rice breeding linesdiscovered 29 SNPs out of which 16 were on intronic region and 13 functional SNPs on exonic region. Out of 13 functional SNPs we obtained, 31% were Synonymous and 45% were Non Synonymous. The latter are beneficial to our study as they result in variation among the species

CHAPTER 4

Conclusion:

An SNP study was performed in order to understand genetic variation which affects the Amylose content in Oryza sativa strains. For the present study, we procured different wild rice varieties from NBPGR, Shimla. Potassium Iodide Test was performed for quantification of Amylose. 17 varieties were chosen and categorized on the basis of Amylose Content. Primers for GBSS-1 and SBE were designed as they majorly contribute for Starch metabolism. 100% amplification was observed for GBSS-1 primers. Sequencing was performed to validate the amplified region and to further perform SNP analysis and structure modeling. In silico SNP Analysis was done using SNPiPlay web based tool. Sequencing GBSS 1 gene involved in starch synthesis in a population of 17 rice breeding lines discovered 29 SNPs out of which 16 were on intronic region and 13 functional SNPs on exonic region. Out of 13 functional SNPs we obtained, 31% were Synonymous and 45% were Non Synonymous. The presence of Non Synonymous SNP's may lead to variation with respect to starch metabolism. These results will further utilized for Association analysis with amylose content. Sequencing results will be used for modelling the GBSS-1 structure and similar study would be performed on SBE so as to understand the structural differences and its effect on starch metabolism.

Future Plan:

- Validation of SNPs
- Association Mapping
- Structure Modeling
- Similar studies for SBE gene.

References:

1. http://www.genoscope.cns.fr/externe/English/Projets/Projet_CC/organisme_CC.html
2. http://www.plantgdb.org/OsGDB/
3. http://btn.genomics.org.cn:8080/rice/
4. http://www.irri.org/about/faq.asp
5. Revedin, A.; Aranguren, B.; Becattini, R.; Longo, L.; Marconi, E.; Lippi, M. M.; Skakun, N.;Sinitsyn,A.eta(2010).)ADP-glucosepryophosphorylase;PPi, pyrophosphate.doi:10.1371/journal.pone.0018385.g001
6. Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (Oryza sativa L.) Manish K. Pandey1, 2,*, †, N. Shobha Rani1, M. Sheshu Madhav1, R. M. Sundaram1, G. S.
7. Varaprasad1, A. K. P. Sivaranjini1, Abhishek Bohra2, G. Ram Kumar1 and Anirudh KumarAluko G, Maetinez C, Tohme J, Castano C, Bergman C, Oard JH. QTL mapping of grainquality traits from inter-specific cross Oryza sativa × O. glaberrima. TheorApplGenet 2004; 109:630-639.
8. Amaravathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, et al. Mapping ofquantitative trait loci for basmati quality traits in rice. Mol Breed 2008; 21:49-65.
9. Asaoka M, Okuno K, Yano M, Fuwa H. Effects of shrunken and other mutations on theproperties of rice endosperm starch. Starch/Starke 1993; 45:383-387.
10. Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD. Microsatellites andsingle nucleotide polymorphism differentiate apparent amylose classes in an extendedpedigree of US rice germplasm. TheorAppl Genet 1997; 94:773-781.
11. Bao JS, Sun M, Corke M. Analysis of genetic behavior of some starch properties in Indicarice (Oryza sativa L): Thermal properties, gel texture, swelling value. TheorApplGenet 2002; 104:408-413.
12. Bao JS, Corke H, Sun M. Nucleotide diversity in starch synthase IIa and variability of singlenucleotide polymorphisms in relation to starch gelatinization temperature and otherphysicochemical properties in rice. TheorAppl Genet 2006; 113:10.1007/S/00122-006-0355-6.
13. Bao JS, Sun M, Zhu LH, Corke H. Analysis of quantitative trait locus for same starchproperties in rice thermal properties, gel texture, swelling volume. J Cereal Sci 2004;39:379-385.

14. SNP in starch biosynthesis genesassociated with nutritional and functionalproperties of riceArdashir Kharabian-Masouleh1, Daniel L. E. Waters1, Russell F. Reinke2,3, Rachelle Ward4& Robert J. Henry5Philpot, K., Martin, M., ButardoJr, V., Willoughby, D. & Fitzgerald, M.Environmental factors that affect the ability of amylose to contribute toretrogradation in gels made from rice flour. J Agric Food Chem 54, 5182–5190(2006).
15. Hu, P., Zhao, H., Duan, Z., Linlin, Z. & Wu, D. Starch digestibility and theestimated glycemic score of different types of rice differing in amylose contents.J Cereal Sci 40, 231–237 (2004).
16. Chung, H. J., Lim, H. S. & Lim, S. T. Effect of partial gelatinization andretrogradation on the enzymatic digestion of waxy rice starch. J Cereal Sci 43,353–359 (2006).
17. Chen, M. H., Bergman, C., Pinson, S. &Fjellstrom, R. Waxy gene haplotypes:Associations with apparent amylose content and the effect by the environment inan international rice germplasm collection. J Cereal Sci 47, 536–545 (2008).
18. Effect of cooking on amylose content of riceAshish Jain, Sughosh M Rao, SarikaSethi, Abhinav Ramesh, SwapnilTiwari, SanjeebKumarMandal, Nishantkumar Singh, AshishSinghal, NaisargModi, VibhavBansalandChitraKalaichelvani European Journal of Experimental Biology, 2012, 2 (2):385-388
19. DJ Jenkins et al. (1981). Am J ClinNutr 34; 362–366.
20. Knowler WC, Barrett-Connor E, Fowler SE, et al., N Engl J Med. 2002;346(6):393–403.
21. Brouns et al. (2005). Nutrition Research Reviews 18; 145–171.
22. Chiu CJ, Liu S, Willett WC, Wolever TM, Brand-Miller JC, Barclay AW, Taylor A., Nutr Rev., 2011 ;69(4):231-42.
23. Temelkova-Kurktschiev TS, Koehler C, Henkel E, Leonhardt W, Fuecker K, Hanefeld M., Diabetes Care. 2000Dec;23(12):1830-4.
24. White Rice, Brown Rice, and Risk of Type 2 Diabetes in US Men and Women."Arch Intern Med 2010 170: 961-969.
25. International table of glycemic index and glycemic load values: 2002."Am J ClinNutr 2002 76: 5-56
26. Brand-Miller JC, Pang E, Bramal L. Am J ClinNutr 1992;56:1034–6.
27. Srisawas, W. and Jindal, V. K. (2007), Journal of texture studies, 38: 21–41. doi: 10.1111/j.1745-4603.2007.00084.x
28. Efficient measurement of amylose content in cereal grainsGongshe Hu*, Charlotte Burton, Chunyan Yang 1 Journal of Cereal Science 51 (2010) 35–40

(VI)

Publications of the Student

1. VarunJaiswal, Sree K Chanumolu, **Pankaj Sharma**, Rajinder S Chauhan and Chittaranjan Rout* EpiCombFlu:(Feb,2013) Exploring known influenza epitopes and their combination to design universal influenza vaccine.OXFORD BIOINFORMATICS : BIOINF 13-0282.R1 Application Note(Communicated 2013)